

Against the Grain

Volume 19 | Issue 1

Article 9

February 2007

Digital Collections, the Next Generation: Transitioning to METS for Science Fiction Digitization Project

Jen Wolfe

University of Iowa Libraries, jennifer-wolfe@uiowa.edu

Mark F. Anderson

University of Iowa Libraries, mark-anderson@uiowa.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

Recommended Citation

Wolfe, Jen and Anderson, Mark F. (2007) "Digital Collections, the Next Generation: Transitioning to METS for Science Fiction Digitization Project," *Against the Grain*: Vol. 19: Iss. 1, Article 9.
DOI: <https://doi.org/10.7771/2380-176X.5230>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Expand Your Library Beyond the Bookshelves



The R2 Library offers a unique business model for digital content enabling you to select and purchase only the resources you need through a topic-based, navigable, highly searchable database. Build your digital collections one resource at a time, as you want it when you want it!

Visit www.rittenhouse today to learn more about the R2 Library and to sign up for a FREE 30- day trial.



A world of health sciences information at your fingertips

511 Fehleley Drive King of Prussia, PA 19406
www.rittenhouse.com 800.345.6425

Digital Collections, the Next Generation: Transitioning to METS for a Science Fiction Digitization Project

by **Jen Wolfe** (Metadata Librarian, University of Iowa Libraries) <jennifer-wolfe@uiowa.edu>

and **Mark F. Anderson** (Digital Initiatives Librarian, University of Iowa Libraries) <mark-anderson@uiowa.edu>

Those of you who use eBay may know the mixture of elation and remorse that can set in after winning an auction for a big-ticket item, and just how quickly the reaction “Hooray! I won!” can be replaced by “Oh no, I won.” Staff at the **University of Iowa Libraries** got a taste of this in 2005, when **Rob Latham**, a professor in the UI’s Department of English and a noted scholar of science fiction, proposed the acquisition of 250,000 science fiction fanzines up for bid in an eBay auction. Quick reactions and skillful negotiations on the part of **Latham** and the Libraries’ Collection Development staff resulted in a withdrawal of the online auction. The owner of the fanzines, science fiction fan and collector **M. Horvat**, happily agreed to a direct sale so that the materials could be kept together and made available to the public in the Libraries’ Special Collections Department.

Even absent any last-minute bidding frenzy, library staff still experienced a form of buyer’s remorse, but not upon purchasing the collection. Instead, that moment occurred several months later when a semi-trailer containing

hundreds of fanzine-laden cardboard boxes arrived on the Libraries’ loading dock. It was then that the reality of the situation could no longer be denied. In order to be a useful tool for scholarly research, somehow this collection of a quarter of a million obscure publications would have to be cataloged.

Like the astronauts pictured spacewalking their way across the fanzines’ covers, the UI’s newly formed Digital Library Services stepped into the void, exploring strange new worlds of hierarchical textual material, digital asset management systems, full-text searching, and copyright navigation — and boldly going where the Libraries hadn’t gone before: cataloging with METS, the **Metadata Encoding Transmission Standard**.

A Major Player in Science Fiction Research

Aside from the not-so-minor problem of bibliographic control, acquiring the **M. Horvat Collection of Science Fiction Fanzines** was a coup for the library. **Professor Latham** stated to the press that the collection made the

University of Iowa a major player in science fiction research, a sentiment echoed by **Sid Huttner**, head of the UI’s Special Collections Department: “We’ve long collected the scholarly literature related to science fiction. As a result of this acquisition, however, we now have the largest collection of science fiction fanzines in the Midwest and one of the largest that exists anywhere.”¹

But who cares, outside the genre’s core audience (i.e., those fluent in Klingon and Wookie, owners of custom-made Imperial stormtrooper outfits)? Plenty of scholars, as it turns out. Like much detritus of popular culture, originally dismissed by critics (and acquisitions librarians) as unworthy of notice, these materials have undergone a reassessment. Indeed, science fiction in general, once thought of as kids’ stuff, is now recognized not only as a vision of possible futures, but also as a vehicle for commentary on contemporary society. According to **Gary Westfahl**, coordinator of English programs at the **University of California-Riverside** and the author and

continued on page 38

editor of 15 books of science-fiction criticism, "the literature of science fiction is now considered worth examining by a broad range of literary scholars, in fields such as feminism, postmodernism, and postcolonialism, as well as experts in other disciplines, such as history and political science."²

Science fiction fanzines — usually low-budget, low-circulation, do-it-yourself magazines created by fans — are particularly valuable for documenting the close relationship between the genre's authors and its audience. In these mimeographed or photocopied pages, leading novelists wrote candid letters to the editor and gave unguarded interviews discussing their works, writing processes, and publishers, while many of the fanzine writers and editors later went on to publish successful novels of their own. As a result, science fiction fanzines are a goldmine of rare primary source material. "When it is fully available to scholars," explains Latham, "the Horvat collection will provide an invaluable resource for research into the history of modern science fiction and the formation of fan communities." [emphasis most emphatically added] (1)

The magnitude of the task of providing useful bibliographic access quickly became apparent. Horvat's fanzines had languished in a storage warehouse for many years, unused and mostly unorganized. The small amount of staff time that Special Collections staff could afford to devote to the materials went towards the most minimal level of processing — sorting, de-duplicating, rehousing, and compiling a comprehensive title list for a finding aid — a project that's only nearing completion almost two years later. As far as traditional cataloging in MARC format, that approach wasn't even considered. The qualities that made the collection valuable — its relative comprehensiveness, the rarity of the materials and their nonconformance to standard publications — rendered it impractical for serials cataloging treatment, let alone the type of item-level treatment with subject access



that would make it truly useful. Although the fanzine collection was open to researchers, the only way to find content on a specific subject involved a laborious process of digging through box after box, picking up individual issues and skimming the pages to see if they contained any relevant information.

Boldly Going

Elsewhere in the Libraries, a new department was being formed to centralize and coordinate digitization efforts both within the library and throughout the campus. The UI's Digital Library Services debuted in January 2006, staffed by a program coordinator, a

against the grain people profile

Jen Wolfe

Metadata Librarian
University of Iowa Libraries
<jennifer-wolfe@uiowa.edu>

BORN & LIVED: Iowa, Chicago, St. Louis, Iowa, Seattle, Chicago, Iowa.

EDUCATION: B.A. in English, **University of Iowa**, 1999; M.L.I.S., **University of Washington**, 2004.

PROFESSIONAL CAREER AND ACTIVITIES: **Experience Music Project**, 2000-2003: digital collections cataloger for rock and roll museum.

University of Iowa Libraries, 2004- : cataloger in tech services, then metadata librarian in digital library services.

Blog coordinator for **ALA ALCTS Networked Resources and Metadata Interest Group**. Currently organizing bake sales and raffles to raise the \$45,000 fee to buy my library's way into the **Digital Library Federation**.

FAVORITE BOOKS: **Jane Austen**; mid-20th century young adult fiction about library careers (see <http://www.jenw.org/home.htm>); **Kelly Link**.

PET PEEVES/WHAT MAKES ME MAD: Reactionaryism, negativity; lack of vegetarian options.

PHILOSOPHY: Speed and efficiency trumps fullness of detail.

HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS:

Further down the current path: shift in focus from creating MARC records for published materials to creating metadata records (of the non-MARC variety) for unique materials (special collections, institutional repository); **Worldcat** instead of local catalogs; automatic generation of metadata; incorporation of **Web 2.0** functionalities (user tags, ratings, recommendations); pressure from patrons for electronic access to everything. 🐼



digital initiatives librarian, and a metadata librarian; the positions were filled with reassignments from the Libraries' Information Technology, Preservation, and Technical Services departments. Seeking high-profile projects with which to make a name for itself, DLS selected the fanzine collection as a prime candidate for digitization.

In addition to the criteria described above, the materials held particular appeal since there was little chance of duplicating the efforts of another institution; according to the **University's** subject experts, none of the few libraries with comparable collections had digitized their holdings. Furthermore, it was unlikely that the local digital collection at the **UI** would be made redundant by mass digitization initiatives such as **Google Books**, a factor that has quickly become a key consideration when working with published materials. An equally important reason for selection was the active use of the fanzines by faculty and graduate students, directing DLS efforts to a "just in time" rather than a "just in case" model. Such collaborations with academic departments support a primary strategic goal of both DLS and of the **UI Libraries** as a whole. As more and more students shun print information resources for those online, libraries wishing to remain a vital force on campus must find innovative ap-

proaches to serving their primary users.

Consideration of copyright restrictions has ruled out, for the time being, the digitization of complete issues of the fanzines, so DLS staff met with the faculty and graduate students conducting research with the materials to decide how best to make the collection more usable to science fiction scholars. They determined that any means of pinpointing search efforts would be most helpful, and that breadth was more valuable than depth — i.e., digitizing a larger number of items using a minimal approach was preferable to digitizing a smaller number using a fuller approach. The goal was to create a searchable database that would serve as an OPAC for the physical collection, saving time on the part of local researchers, saving wear and tear on the artifacts (and on Special Collections staff), and allowing at least partial access to the collection for a much broader audience. Based on this assessment and on consultations with Special Collections staff who were most familiar with the materials, DLS developed a plan for a science fiction fanzine digital collection that on paper appeared to comprise four simple steps:

- full color scanning of the covers and table of contents pages
- OCR (optical character recognition)

continued on page 39

<<http://www.against-the-grain.com>>

- scanning of the TOC images in order to create a text file for keyword searching
- minimal item-level metadata cataloging
- combining these elements (image files, text files and metadata records) into digital objects and uploading them to a digital asset management system

Unfortunately, things didn't go quite as easily as planned.

A Useful Standard

Even before the formal establishment of Digital Library Services, its staff members had acquired significant experience building digital collections as part of the Libraries' consortial participation in Iowa's statewide digital cultural heritage initiative. The UI's contributions chiefly consist of digitized image collections (e.g., historic photographs) and simple textual materials (e.g., newspaper clippings); these are cataloged using the **Dublin Core** metadata element set, and uploaded to the asset management system **CONTENTdm**, developed by **DiMeMa** and recently acquired by **OCLC**.

For the fanzines, DLS decided to adopt a different workflow due to two factors. First, both faculty and library staff hope eventually to encourage other libraries with science fiction fanzine collections to participate in a consortia, filling gaps in the **Horvat** digital collection with digitized holdings of their own, resulting in a much more comprehensive and useful tool for scholars. To enable this goal, following metadata standards as much as possible in order to create shareable records is of primary importance. Second, the more complex hierarchical physical structure of a text such as a book or a fanzine — with a series of pages meant to be displayed in a particular order — requires a correspondingly more complex digital object, with structural metadata reflecting the item's organization. (Although the current plan for the fanzines requires less complexity, since only the cover and the table of contents page are being digitized, DLS wished to create a digital collection that would leave open the possibility of full-text digitization in the future.)

While **CONTENTdm** can accommodate compound objects such as fanzines, its method of organizing and displaying the digitized pages is a proprietary function of the software, resulting in nonstandard structural metadata records that can't easily be migrated from one system to another. However, the **UI Libraries** had recently acquired an additional digital asset management system, the **ExLibris** product **DigiTool**, and had been seeking an opportunity to try it out. Although the software was less familiar to staff than **CONTENTdm**, **DigiTool** was selected for the project, since it accommodates a wider variety of standards — including **METS**, the **Metadata Encoding Transmission Standard**.

Developed by the **Library of Congress** as a standard for enhancing the capability for data migration and the long-term durability of

complex digital objects, the **METS** schema provides:

a flexible mechanism for encoding descriptive, administrative, and structural metadata for a digital library object, and for expressing the complex links between these various forms of metadata. It can therefore provide a useful standard for the exchange of digital library objects between repositories.³

Often described as a "wrapper," **METS** isn't an alternative to **Dublin Core**, but rather a schema for creating a standardized container that can include **DC** records along with additional data necessary for the internal management of a digital object. The schema consists of seven major sections: **METS** header, descriptive metadata, administrative metadata, file section, structural map, structural links, and behavior.

As **METS** neophytes, DLS staff designed a simplified exemplar for the fanzines, incorporating administrative metadata into the descriptive metadata section, and omitting the structural links and behavior sections, which were deemed unnecessary for the project. Modeling records after the samples provided both by **DigiTool** and on **LC's METS** site proved fairly straightforward; for each item in the fanzine collection, the **METS** wrapper provided a consistent order for the digital object thumbnail, image files for cover and table of contents pages, TOC text file, and metadata record. The most time-consuming part of the process was also the most familiar — choosing data elements and syntax for the descriptive metadata section.

Since quantity had been prioritized over fullness and detail, a fairly simple record tem-

continued on page 40

Mark F. Anderson

against the grain people profile

Digital Initiatives Librarian
University of Iowa Libraries
<mark-anderson@uiowa.edu>

BORN & LIVED: Born in Ames, Iowa. Grew up there and attended college close by at **Iowa State University**.

EARLY LIFE: Immersed in music from an early age. I've had the opportunity to sing in a concert with the **Chieftains** and play trumpet with **Marvin Hamlisch**. Also, many family trips developed in me a love of history and geography.

FAMILY: Married to **Kristin**, a middle school music teacher. We have 2 cats.

EDUCATION: B.A. (Music), **Iowa State University**, 2002. M.A. (Library and Information Science), **University of Iowa**, 2004.

FIRST JOB: First paying job, daily cleaning of law office during high school. First library job, work study copy cataloging for music scores and CDs. First professional job, Digital Images Cataloger for the **Father Catich Digital Collection**, **St. Ambrose University**. Assisted with the **Iowa Heritage Digital Collections** statewide cultural heritage consortium. Now, Digital Initiatives Librarian at **University of Iowa Libraries**.

IN MY SPARE TIME I LIKE TO: Go on long bicycle rides and photograph bald eagles, which are making a comeback on Iowa rivers.

FAVORITE BOOKS: Currently reading *Jitterbug Perfume* by **Tom Robbins**. Just read *The Long Emergency* by **James Howard Kunstler**. I like anything by **Neal Stephenson**.

PET PEEVES/WHAT MAKES ME MAD: Bad attitudes.

PHILOSOPHY: How goals are accomplished is usually more important than what goals are accomplished.

MOST MEANINGFUL CAREER ACHIEVEMENT: Helped digitize and make available through the digital library an important series of local history books that means a great deal to the community.

GOAL I HOPE TO ACHIEVE FIVE YEARS FROM NOW: I hope to see a successful program of mass digitization at our library.

HOW/WHERE DO I SEE THE INDUSTRY IN FIVE YEARS: More interactive services and opportunities for library users to interact with each other online. But libraries will still be important gatekeepers and filterers of information. I hate to make too many bold predictions that can be used against me when it all goes differently! 🐼



plate was designed, based on qualified **Dublin Core**. Catalogers would capture basic information such as fanzine title, number, date, and extent. Except for the defaulted LCSH string "Science fiction — Periodicals," no attempt would be made to provide controlled subject access. This decision was based not only on the need for expediency due to the collection's size, but also on the narrow subject range of the materials and the lack of subject expertise among the metadata catalogers, as well as the inclusion of keyword searchable text from the table of contents pages. One exception was made for major science fiction authors, who would receive subject access in the form of authorized **LC** headings; this would allow for easier collocation when performing federated searching of other databases such as the Libraries' **OPAC**.

The Collection Strikes Back

Perhaps unsurprisingly, the new department encountered a variety of problems while executing this plan to simultaneously break in a new material format, a new metadata schema, and a new digital asset management system. While it all looked good on paper, the reality provided many opportunities to develop problem-solving skills and flexibility, two essential components of any digitization project.

- **Materials:** the **Horvat** fanzines, spanning the 1940s through the 1990s, range in format from primitive, pre-mimeograph newsletters to offset-printed, magazine-like publications with full-color covers and ISSNs. The lack of standardization in the materials caused no small amount of consternation on the part of metadata catalogers. Some fanzines had no table of contents, some had no cover, some had neither. Issue enumeration was similarly inconsistent, with the labels "volume," "number," and "issue" occasionally used interchangeably; for titles with longer press runs, the numbering was subject to starting over without notice, resulting in non-unique identifiers. Title changes abounded, with fanzines like *Algol* becoming *Algol Starship*, then later just *Starship*; such difficulties, which may sound all too familiar to serials librarians, were more aggravated with the fanzines due to their amateur publication status. In the interest of moving the project forward, catalogers were urged to adopt a pragmatic approach — to do the best they could at creating consistent records without trying to get them perfect. Unlike traditional cataloging for the library **OPAC**, accuracy wasn't as important since possible discrepancies in the data could be checked by the user looking at the page images onscreen.

The OCR treatment of table of contents pages was similarly problematic, often requiring extensive editing due to the low-quality printing of the type, darkened pages brittle with age, and unusual layouts and fonts. For

some titles, catalogers found it more efficient to just type the data from scratch, rather than attempting to edit the scanned text.

- **Metadata:** two major adjustments were made to the original workflow plans for **METS** record creation. First, the application of authorized subject headings for prominent writers was postponed; attempts to do so caused a significant slow-down in the process, especially since catalogers lacked sufficient subject expertise to quickly recognize important names. Library staff decided it would be more effective and efficient to work towards providing this subject access as a separate step in the process, possibly performed by the English departments' graduate student assistants.

XML editing of the **METS** files also proved challenging, since the records had to be created outside of and then uploaded to the digital asset management system. To solve this problem, DLS staff worked with the Libraries' Web team to create a Web-based form that hid much of the **METS** record from view of the cataloger, displaying only the values that required changing, which greatly simplified the data entry process.

- **Digital asset management system:** adjusting to the **DigiTool** workflow was probably the most difficult part of the project. **CONTENTdm** puts much more control in the hands of digital librarians; setting up collections and changing their metadata fields is a fairly trivial process. Using **DigiTool**, on the other hand, requires much more assistance on the part of systems librarians and programmers. Furthermore, while **CONTENTdm** is a well-developed, widely-adopted system, **DigiTool** definitely had some bugs that needed to be worked out.

As one of the first institutions to implement the program's most recent release, the **University of Iowa** spent quite a bit of time working with developers at **ExLibris** about certain issues with the software. Some were easily fixed; for example, **DigiTool**'s metadata editing module allowed editing of the descriptive metadata section of the **METS** file, but wouldn't save records if they had more than one Creator field present. DLS staff explained that unlike the **MARC** records used in **ExLibris**' automated

library system product, digital library metadata records weren't limited to only one main entry creator field, so the developers immediately fixed the problem. Other less trivial difficulties, such as the inability to upload **METS** records in batches, are still being worked on. Such inconveniences were balanced by the company's responsiveness and commitment to standards, which supported the Libraries' long-term goals for the fanzine project.

A New Hope

With the **UI's Science Fiction Fanzine Digital Collection** up and running, library staff have been taking steps to build on its success. Special Collections librarians are investigating grant funding to organize a science fiction fanzine symposium that could eventually lead to the establishment of a related library consortium, while DLS staff are looking into mainstreaming the project's workflow, shifting metadata cataloging duties from the department's student assistants to paraprofessional staff in the Libraries' technical services department. But the most valuable takeaways from this experience may not be just a new collection and a wider repertoire of metadata standards. Perhaps more importantly, staff members learned that building an innovative digital collection — taking on new challenges, addressing user needs, and refusing to be limited by technology — doesn't have to be science fiction.

References

1. **University News Service.** 2005. UI Libraries acquire huge collection of science fiction fanzines. University of Iowa news release, February 11, 2005. <http://www.news-releases.uiowa.edu/2005/february/021105zines.html> (accessed January 10, 2007).
2. **Matson, R.** 2006. Especially collected. *The Daily Iowan*, July 27, 2006. <http://www.dailyiowan.com/media/storage/paper599/news/2006/07/27/80Hours/Especially.Collectd-2134631.shtml> (accessed January 10, 2007).
3. **Library of Congress.** 2006. **METS: an overview & tutorial.** **METS: Metadata Encoding & Transmission Standard Official Website**, September 13, 2006. <http://www.loc.gov/standards/mets/METSOverview.v2.html> (accessed January 10, 2007). 🐾

Rumors

from page 36

article by **Joe Bolger** — "Reed Elsevier, the Anglo-Dutch media group, has kick-started plans to sell off its education division, as it seeks to focus its investment on its three other divisions. ... "The announcement came as the group reported a dip in **Harcourt**'s fortunes amid declining textbook markets in the US. The division reported a 20 percent drop in operating profits to 129M pounds for last year, with US textbook markets estimated to have

weakened by 6 percent in 2006." business.timesonline.co.uk/tol/business/industry_sectors/media/article1388383.ece

Held on **February 8th in London**, a range of leading authors, researchers, funders, librarians and publishers came together at a recent **BioMed Central** colloquium entitled "**Open Access: How Can We Achieve Quality and Quantity?**" Featured speakers include: **Sir Muir Gray** (Director of Clinical Knowledge, Process and Safety for England's National Health Service Connecting for Health); **Rob-**

continued on page 52